Epigenetic Methodology: Focus on DNA Methylation

NIEHS/EPA CEHC Meeting

Epigenetics Breakout Session

Jaclyn Goodrich and Dana Dolinoy

Univ. of Michigan School of Public Health

Formative Children's Environmental Health and Disease Prevention Research Center

(P20 ES018171/RD834800, PI: Karen Peterson)

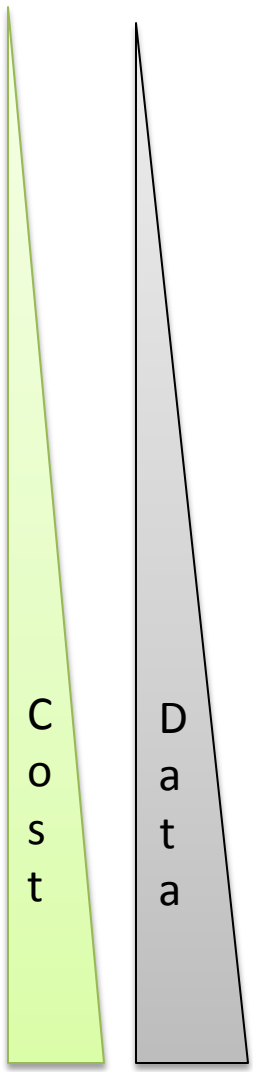
M SPH

Epigenetics in Children's Environmental Health

- Why?
  - Potential mechanism linking early life exposures to adult disease
- Considerations
  - DNA quantity and quality
  - DNA source
  - Available equipment
  - BUDGET
- How?
  - Multiple methods exist to assess DNA methylation, histone modifications, non-coding RNA
  - Outline major methods for analyzing DNA methylation

Three Basic Layers of Analyses

General Trend:

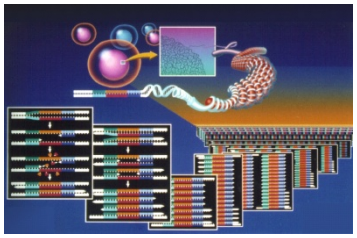


Global
- Repetitive elements → LINE, Alu
- Restriction enzyme based → LUMA

Gene Specific
- Bisulfite Conversion → MS PCR
- Bisulfite Conversion → Pyrosequencing
- Bisulfite Conversion → Sequenom

Epigenome-Wide
- Biased → Arrays
- Unbiased → Parallel Sequencing
- Unbiased → Arrays

Methylation Conundrum



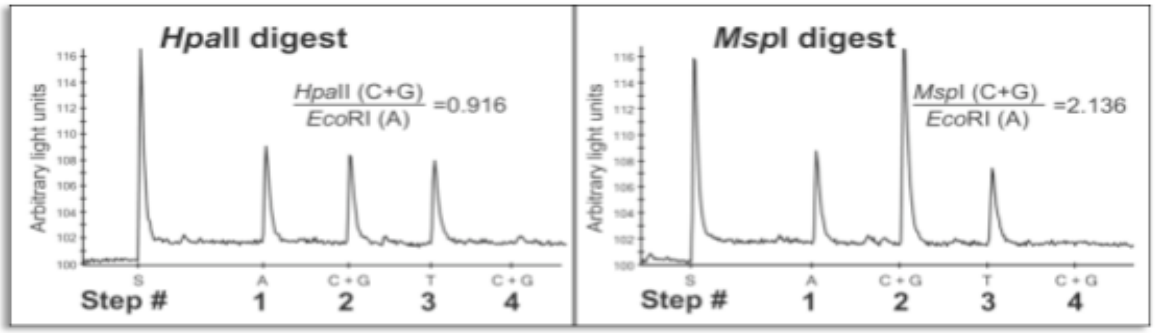
- DNA methylation information is erased by standard molecular biology techniques!
  - Erased by PCR
  - Not revealed by hybridization or sequencing

- Need Methylation-dependent pretreatments
  - Enzyme Digestion, Affinity Enrichment, or Bisulfite Conversion

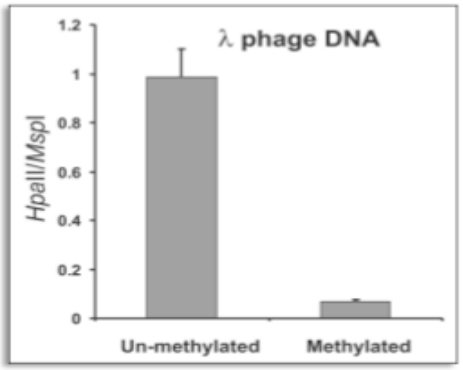
Laird, P. *Nature Review Genetics,* 2010

Global Methylation: Restriction Enzyme-Based Method

- LUminometric Methylation Assay (LUMA) – Determines methylation at all CCGG sites
  - *Pro:* Do not need to know sequence
  - *Con:* Does not give information on which genes are affected



*Hpa*II cuts unmethylated CCGG

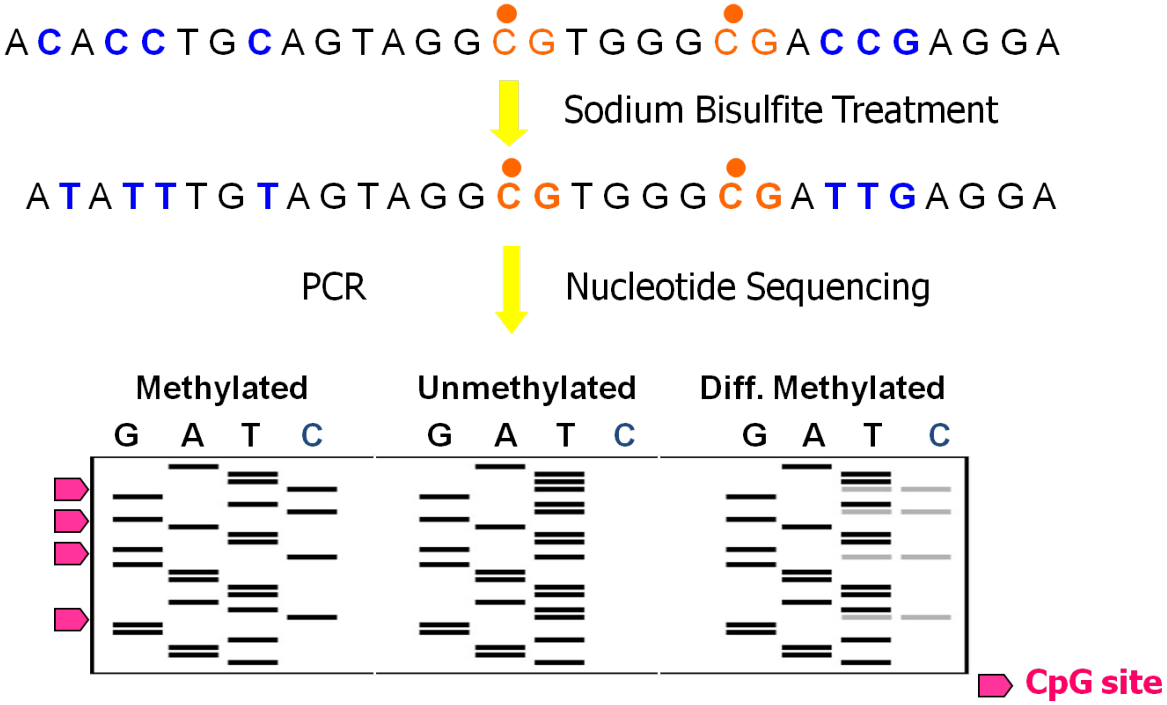


*Msp*I cuts methylated and unmethylated CCGG

Karimi, M. *Epigenetics,* 2006

Pre-Treatment: Bisulfite Conversion

- Treatment necessary for many downstream applications
  - *Pro:* Creates sequence-dependent differences.
  - *Cons:* Harsh treatment with low yield of DNA post-treatment. Creates sequence redundancy- challenge for primer design.



Global Methylation: Repetitive Elements

- LINE1= Long INterspersed Element 1
  - Autonomous retrotransposon representing ~17% of DNA sequence
- Alu= a SINE family (Short INterspersed Element)
  - Non-autonomous retrotransposon representing ~11% of DNA sequence
- Pros
  - Quick screen
  - Quantitative
- Cons
  - Assay design can be tricky (which region/CpGs to include, problems with variability)
  - Sample quality and tissue of origin can impact assay success

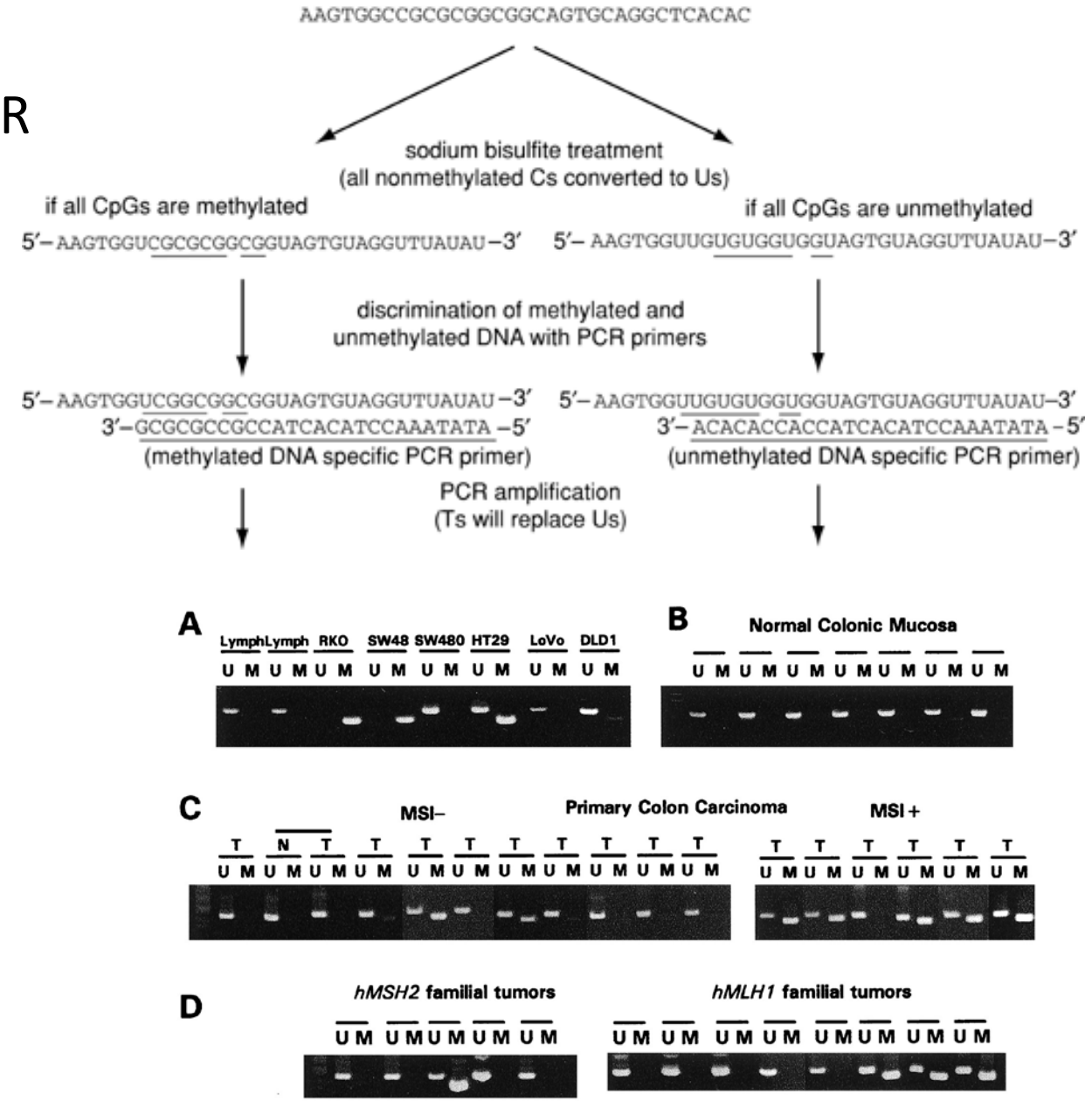
Classes of interspersed repeat in the human genome

			Length	Copy number	Fraction of genome
LINEs	Autonomous	ORF1 ORF2 (pol) AAA	6–8 kb	850,000	21%
SINEs	Non-autonomous	A B AAA	100–300 bp	1,500,000	13%

Lander et al. *Nature,* 2001.

Gene Specific: Methylation-Specific PCR

- Pros
  - Sensitive
  - Quick
  - Inexpensive
- Cons
  - Specificity
  - Not quantitative
  - Only 1-2 CpG sites at a time
  - No bisulfite conversion check

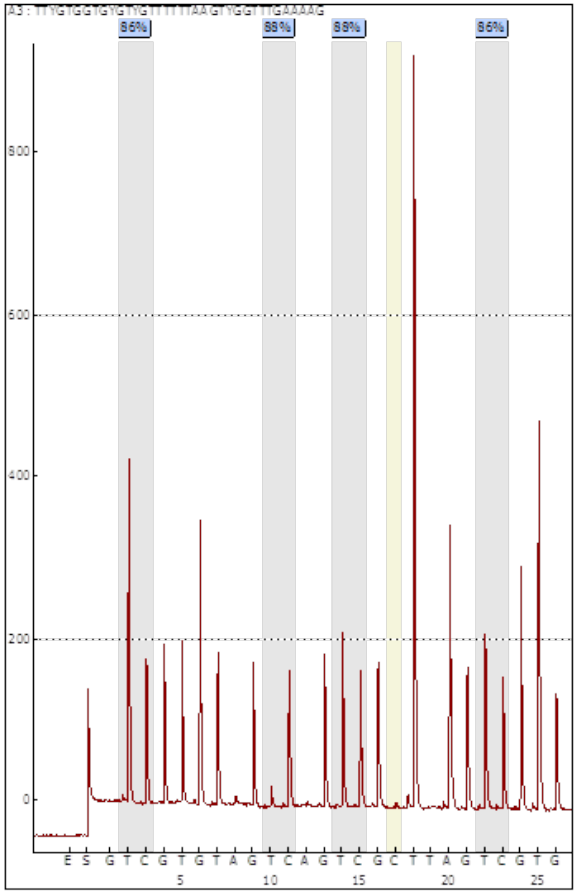


www.currentprotocols.com/ protocol/hg1006; Herman JG et al. *PNAS* 1998; 95:6870-6875

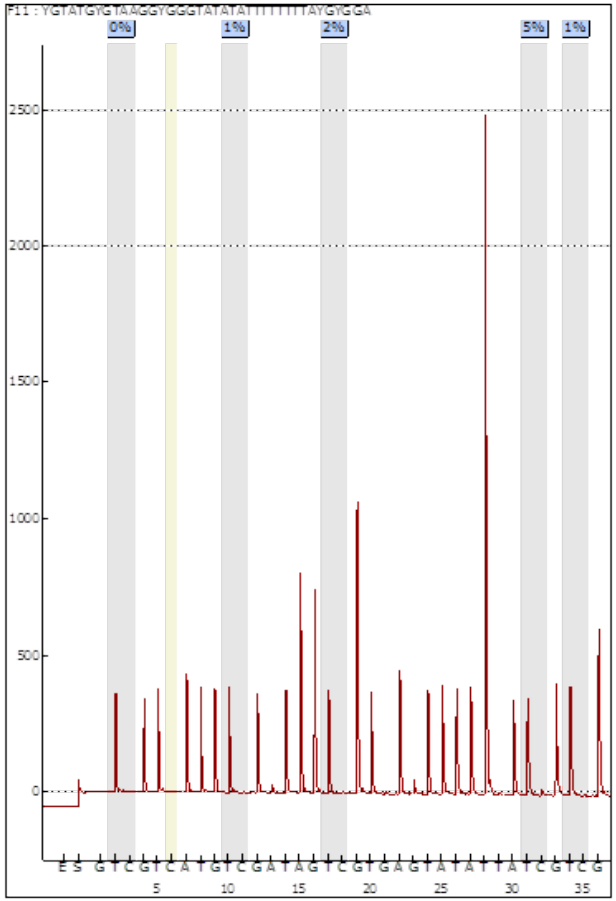
Gene Specific: Pyrosequencing

- Pros
  - Site specific
  - Quantitative
  - High throughput
- Cons
  - Short reads (~100 bp)
  - Assay design difficult in CG dense areas

Heavy Methylation (*LINE*)



Low Methylation (*DNMT1*)

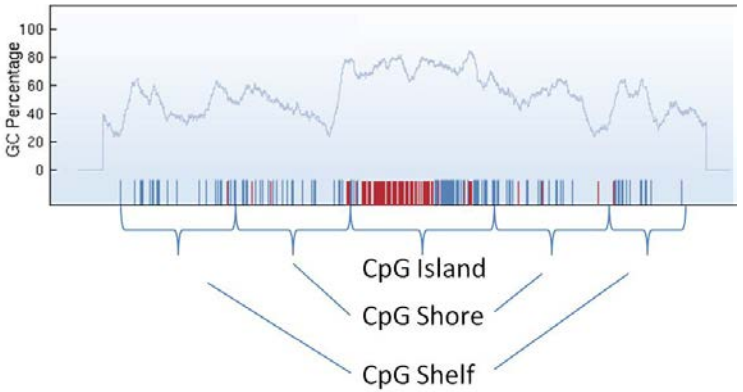


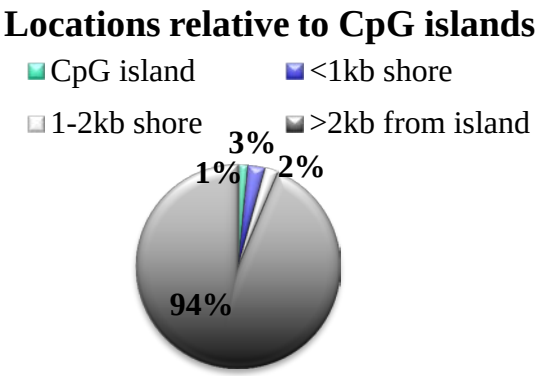
Gene Specific:
Sequenom (Mass Spec System)

- Pros
  - Long reads
  - Inexpensive
- Cons
  - Lose site-specific effects if CpG sites are close together or fall on small fragments

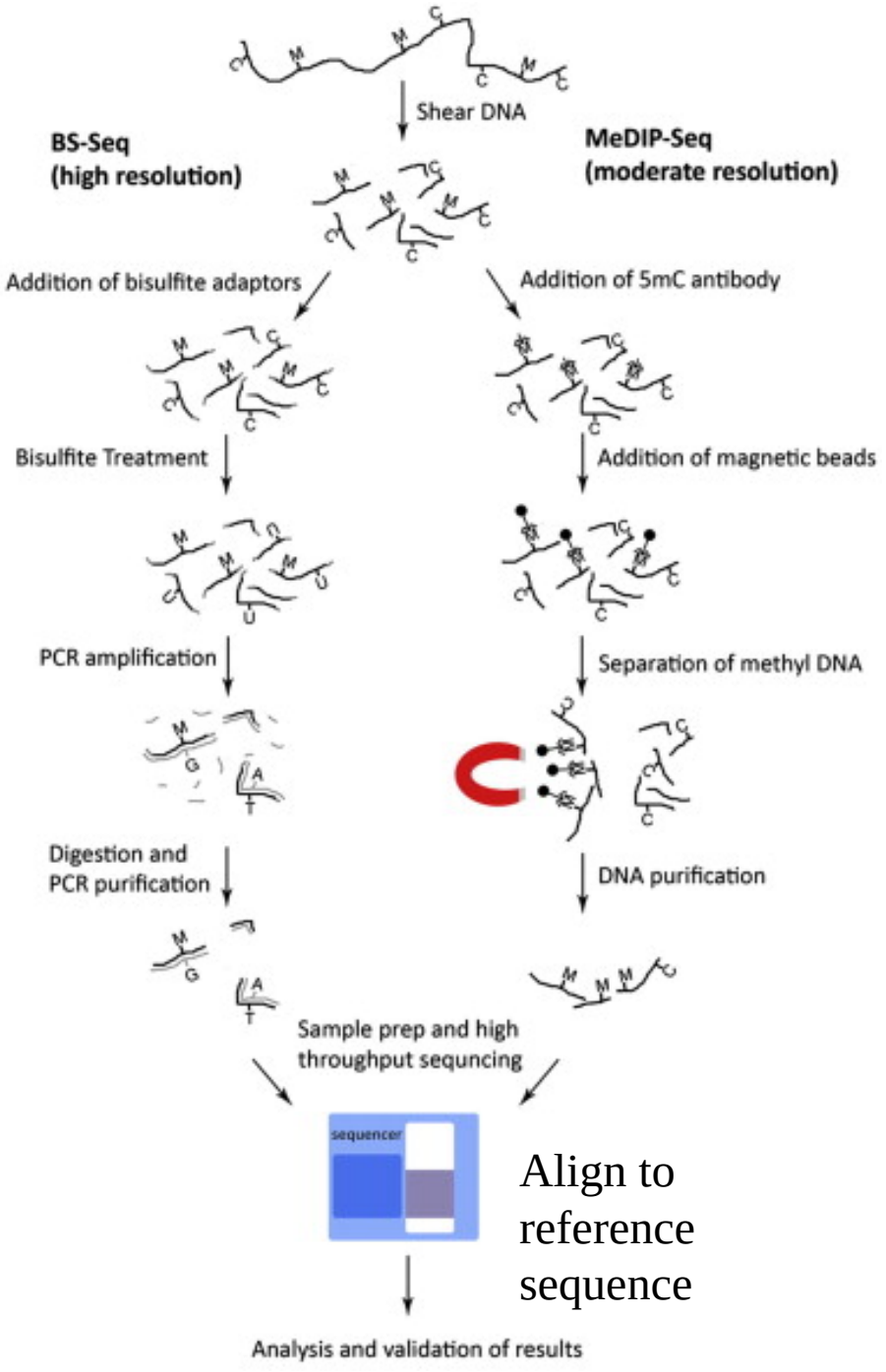
Epigenome-Wide Biased and Unbiased: Arrays

- Illumina BeadChips
  - 27K (CpG Island heavy) vs. 450K (better whole genome coverage)
  - For human DNA
- Promoter Array
  - Biased
  - Available for multiple species
  - Ex: NimbleGen, Agilent
- Tiling Array
  - Affymetrix, NimbleGen, etc.
  - Unbiased
  - For human DNA
  - +41 million probes
- Protocol
  - Antibody precipitation of methylated DNA or MBP before array hybridization





Tiling Arrays Performed with GenPathway (San Diego, CA)



Align to reference sequence

Epigenome-Wide Unbiased: Parallel Sequencing

- Deep Sequencing methods
  - BS-Seq
  - MeDIP-Seq
  - MethylPlex-Seq
- Pros
  - Unbiased whole genome!
- Cons
  - Need reference sequence for BS-Seq method
  - Need antibody for MeDIP method
  - Need to validate findings

Pomraning, Smith, Frietag, *Methods*, 2008

Study Design Considerations (Sample Thru-put)

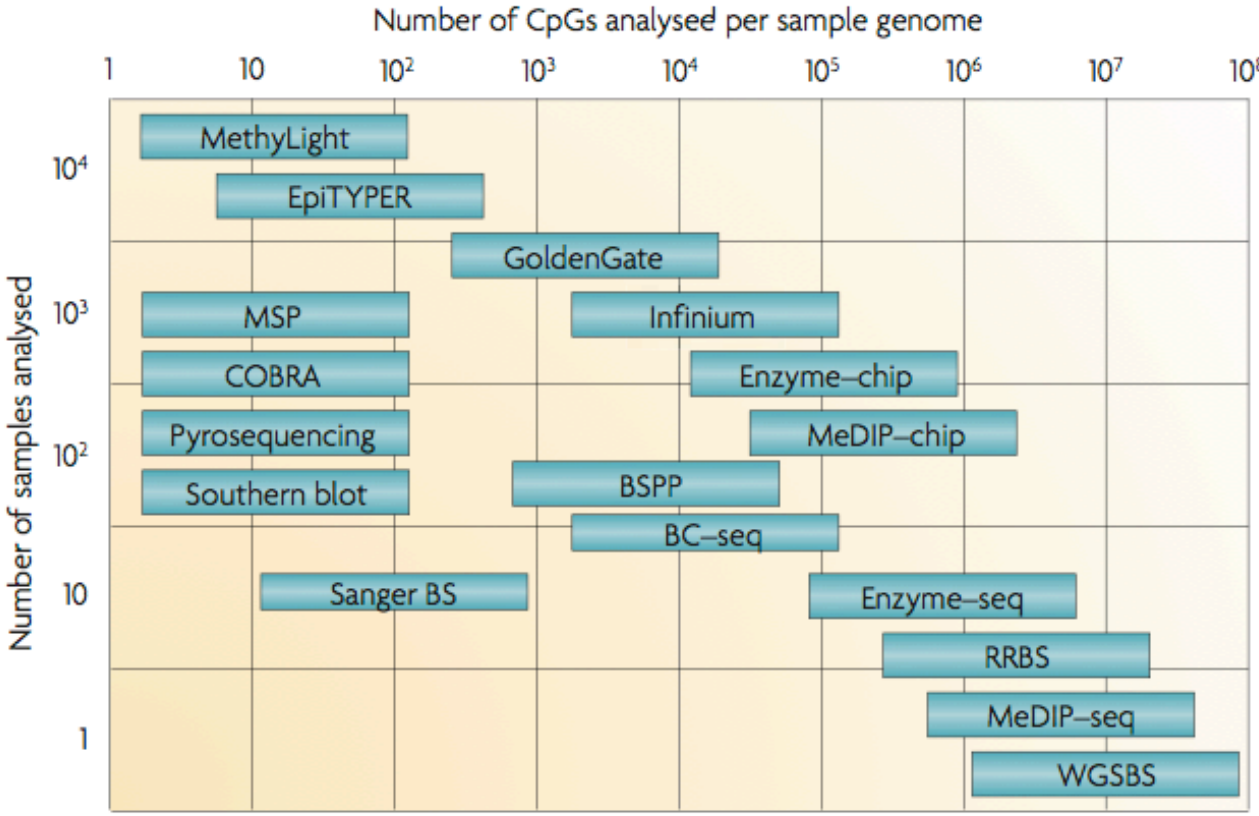


Figure 1 | **Sample throughput versus genome coverage.** A plot of sample throughput against genome coverage for various DNA methylation techniques. Throughput is determined by the number of samples that can be analysed per experiment, based on large eukaryotic genomes. Coverage is determined by the number of CpGs in the genome that can be analysed per experiment. BC–seq, bisulphite conversion followed by capture and sequencing; BS, bisulphite sequencing; BSPP, bisulphite padlock probes; –chip, followed by microarray; COBRA, combined bisulphite restriction analysis; MeDIP, methylated DNA immunoprecipitation; MSP, methylation-specific PCR; RRBS, reduced representation bisulphite sequencing; –seq, followed by sequencing; WGSBS, whole-genome shotgun bisulphite sequencing.

Laird, P. *Nature Review Genetics,* 2010

Study Design Considerations

- Sample requirements (amount and quality)
  - Some restriction digests (>2ug DNA)
  - Illumina can deal with degraded DNA
  - Affinity more tolerable of DNA (im)purity but require large amounts
- Sample throughput
  - High throughput 96 or 384 sample assays are low-labor but high reagent costs
- Genome coverage and resolution
  - Restriction technologies limited to number/distribution of recognition sites
  - Some technologies better for smaller genomes
- Accuracy and reproducibility
  - Fragment length affects hybridization, sequencing, and reproducibility/false positives and negatives
  - Incomplete bisulfite conversion
  - Validation
- Bioinformatics and data storage

Laird, P. *Nature Review Genetics*, 2010